GENOMICS 29, 397–402 (1995)

Human Genomic Characterization of a Novel Locus-Specific Repetitive Sequence

WAN-LIANG LI, CLARK S. HUCKABY, RICHARD E. KOURI, AND GUALBERTO RUANO[1]

BIOS Laboratories, Inc., 5 Science Park, New Haven, Connecticut 06511

Received March 21, 1995; accepted July 18, 1995

A novel human chromosome locus-specific repetitive sequence was identified and characterized using arbitrary PCR. The repeat monomer consensus sequence is 100 bp long, and there are a minimum of 140 to 160 copies of the repetitive sequence per haploid human genome. The repetitive sequence is highly clustered on 20q12 within a 200- to 400-kb region. The highly polymorphic repeat array is inherited in a stable Mendelian fashion. Hybridization analysis revealed detectable conservation of the repeated element only among hominoids and Old World monkeys, where repeat arrangements are also polymorphic. © 1995 Academic Press, Inc.

INTRODUCTION

Repetitive sequences are ubiquitous in both simple and complex genomes. While their biological functions are not fully understood, repetitive sequences have been widely exploited in gene mapping, disease gene localization, and DNA fingerprinting. In the human genome, high-abundance repetitive sequences have been well characterized; the most abundant repeat families are short interspersed elements (SINE, e.g., *Alu*) and long interspersed elements (LINE, e.g., Kpn), which account for roughly 30% of the human genome by mass (Moyzis *et al.*, 1989). Less abundant known repeat elements include THE/*Mst*II sequences (Paulson *et al.*, 1985; Mermer *et al.*, 1987; Fields *et al.*, 1992), minisatellites or variable numbers of tandem repeats (VNTRs) (Jeffreys *et al.*, 1985; Nakamura *et al.*, 1987), microsatellites or short tandem repeats (STRs) (Weber and May, 1989), and telomeric and centromeric alphoid repeats (Moyzis *et al.*, 1987). However, Moyzis *et al.* (1989) calculated that there are up to 1000 different families of repetitive sequences in the human genome yet to be identified. More recently, Kaplan *et al.* (1991) estimated that up to 2.42×10^5 individual repeat elements in various medium reiteration repeat families remain to be discovered.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. U22345.

[1] To whom correspondence should be addressed. Telephone: (203) 773-1450. Fax: (203) 562-9377.

Most of the currently well-characterized human repetitive sequences are distributed in either interspersed or tandem fashion. For example, *Alu* sequences number approximately 10^6 copies in the human genome (Hwu *et al.*, 1986) and occur on average every 3 to 6 kb. Variable CA repeats occur every 40 to 50 kb (Weber and May, 1989). Their random distribution and high variation in tandem length render CA repeats very powerful as genetic markers, which has revolutionized genotyping and disease gene discovery. High abundance, randomly distributed repeated sequences are also used in constructing physical maps (Stallings *et al.*, 1990; Nelson *et al.*, 1991; Lane *et al.*, 1992). Application of low-abundance repetitive sequences in assembling and verifying YAC contigs was also reported (Zucchi and Schlessinger, 1992). Although randomly distributed repetitive sequences are useful tools, there is also a need for locus- or band-specific repetitive sequences for distinguishing individual chromosomes or chromosomal regions. The first identified chromosome-specific repetitive sequences were centromeric α-satellite and satellite-like sequences (Waye and Willard, 1985; Moyzis *et al.*, 1987). Das *et al.* (1987) reported the first chromosome band-specific minisatellite, which was located in the 19q13.3–qter region.

Here we report a new nonalphoid locus-specific repetitive sequence discovered with arbitrarily primed PCR or "RAPD" (random amplified polymorphic DNA, Williams *et al.*, 1990).

MATERIALS AND METHODS

Genomic DNAs. Commercially available genomic DNAs of individual humans, human families, and nonhuman primates, and DNAs from domestic mammals, fruit fly, fish, mussel, lobster, frog, chicken, bacteria, yeast, and nematode were prepared at BIOS Laboratories using established protocols.

RAPD-PCR. RAPD reactions were performed on human DNAs from nine unrelated individuals of the following descent: African (one), Asian (two), and Caucasian (six). Arbitrary primer 5′ ACGACCCACG 3′ at 2.5 μM was used in 25-μl reactions containing 100 ng genomic DNA, 200 μM each dNTP, 1× PCR buffer (BIOS OptiTaq buffer E containing 50 mM KCl, 10 mM Tris–HCl, pH 8.4, 0.01% gelatin, 0.1% Triton X-100, and 3.75 mM $MgCl_2$), and 1.25 units *Taq* DNA polymerase (Perkin–Elmer AmpliTaq). PCR was performed in a Perkin–Elmer 480 thermal cycler using 30 cycles of 1 min each at

0888-7543/95 $12.00

94, 38, and 72°C. The RAPD PCR products were electrophoresed on an ethidium bromide-stained 1.5% agarose gel. Most of the bands exhibited no variation between individuals, but a 730-bp band appeared to be specific to four of the individuals. DNA from this band was recovered from agarose gels using the GeneClean method (BIO 101, Inc.).

Southern blotting and hybridization. Commercially available BIOS blots were used. For RFLP analysis, BIOS RFLP identification blots were used; these blots contain DNA from four unrelated human individuals digested separately with each of nine restriction enzymes (*Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Pst*I, *Pvu*II, *Rsa*I, *Msp*I, and *Taq*I). BIOS Evo blots consist of three different groups: the primate group (cotton-top tamarin, macaque, gorilla, orangutan, chimpanzee, and human), the mammalian group (dog, cat, rabbit, cow, sheep, mouse, rat, hamster, pig, cotton-top tamarin, and human), and the genetic model group (*Escherichia coli*, yeast, nematode, fruit fly, fish, mussel, cow, frog, chicken, mouse, and human). The mass of bacterial, yeast, nematode, and some nonmammalian DNAs was adjusted to compensate for varying genome complexities to ensure similar numbers of genome equivalents. For pedigree analysis, a BIOS family allelotyping blot containing DNAs from a 14-member, three-generation family (CEPH Family 1333) was used.

The probe used for all filter hybridization was the 730-bp PCR product (see Fig. 1). The probe was radiolabeled using the BIOS Tag-It kit (Ruano *et al.*, 1993). Hybridization was performed in 10 ml of a modified Church and Gilbert solution (0.5 *M* sodium phosphate, 1.0 m*M* EDTA, and 7.0% SDS, pH 7.2) at 65°C overnight in a hybridization oven (National Labnet Co.). Washing was carried out at high stringency (0.2× SSC, 0.2% SDS at 65°C), and blots were exposed to X-ray film for 10 min to 3 h.

Chromosomal localization and FISH analysis. Chromosomal localization by Southern blot analysis was carried out by using both monochromosomal and polychromosomal BIOS somatic cell hybrid panels. The probe was mapped cytogenetically by fluorescence *in situ* hybridization (FISH). A cloned PCR product (described below) was labeled with biotin–dUTP by nick-translation, prehybridized with sheared human DNA, and then hybridized to normal human metaphase chromosomes derived from PHA-stimulated peripheral blood lymphocytes in a solution containing 50% formamide, 10% dextran sulfate, and 2× SSC. Specific hybridization signals were detected by incubating the hybridized slides with fluorescein-conjugated avidin. The chromosomes were then counterstained with propidium iodide. Some of the slides were also cohybridized with a biotinylated human chromosome 20-specific centromere probe, D20Z1 (Oncor), to show the centromeric position of chromosome 20. Band assignment was further carried out with fractional length (FL) correlation (Lichter *et al.*, 1990). The distance of the signal from the centromere to the telomere was measured on at least 10 metaphase chromosomes 20. The mean FL value was calculated from these measurements. The mean FL value was then plotted onto the ISCN (International System for Human Cytogenetic Nomenclature) 1985 idiogram of chromosome 20 for band assignment.

Cloning and sequencing. The PCR product was cloned into the plasmid vector pCRII (Invitrogen) using the manufacturer's protocol. Sequencing was performed on both strands using Sequenase 2.0 (USB) according to the manufacturer's protocol for sequencing double-stranded templates. The sequencing primers were 15-mer oligonucleotides (GCCAGTGTGCTGGAA and TGGATATCTGCAGAA for forward strand and reverse strand, respectively) based on the vector sequence flanking the insert.

Copy number estimate. Dot blots were used to estimate the copy number of the repetitive sequence in the human genome. Human genomic DNA amounts of 17, 170, 850, 1700, and 3400 ng corresponding to 5×10^3, 5×10^4, 2.5×10^5, 5×10^5, and 10^6 haploid genome equivalents, respectively, were denatured and immobilized in dots on positively charged nylon filters. Plasmid DNA amounts of 0.026, 0.26, 1.3, 2.6, and 5.2 pg (4700 bp plasmid including the 730-bp insert), which corresponds to 5×10^3, 5×10^4, 2.5×10^5, 5×10^5, and 10^6 copies, respectively, were mixed with 17, 170, 850, 1700, and 3400 ng of mouse DNA, respectively. The mixture was then denatured and immobilized in dots on the same filters. The mouse carrier DNA was included to eliminate the bias resulting from easy access of probe to a simple cloned molecule. Previous hybridization studies showed that no cross-hybridization of the repetitive element with the mouse genome is observed under our hybridization conditions. The filters were then hybridized to labeled repetitive sequence probe, and the hybridization signals in each titration were compared using autoradiography.

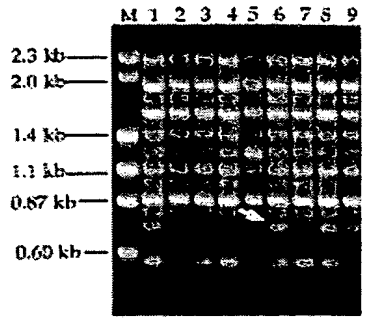

FIG. 1. RAPD fingerprints of nine human individuals (lanes **1** and **3**, Asians; lane **5**, African-American; lanes **2**, **4**, **6**, **7**, **8**, and **9**, Caucasians; lane **M**, molecular weight markers). Amplification conditions are given in the text. The polymorphic band (730 bp, arrow) was confirmed to contain a novel repetitive element.

Genomic clones. An arrayed P1 library representing a 1.2× coverage of human genome was accessed from the Reference Library Database (RLDB; Zehetner and Lehrach, 1994; Francis *et al.*, 1994) through the BIOS P1 library screening service. The 730-bp probe was radiolabeled as described above and hybridized to the high-density filters in 10 ml of modified Church and Gilbert hybridization solution at 65°C overnight following a 2-h prehybridization in the same solution with 100 μg/ml of salmon sperm DNA as blocking agent. After washing as described for Southern blots, the filters were autoradiographed at room temperature for 24 h, and candidate positive clones were identified. Single colonies of candidate positive clones were grown in 100 ml TY broth containing 50 mg/ml kanamycin at 37°C overnight with moderate shaking (about 200 rpm). P1 DNAs were isolated using the Qiagen midiprep protocol for low-copy-number plasmids. Purified P1 DNA with human inserts and P1 vector (pAd10 SacBII) DNA controls were digested with *Hin*dIII, electrophoresed in 1.0% agarose gels, transferred into charged nylon membranes, and hybridized to ^{32}P-labeled 730-bp probe for confirmation.

RESULTS

RFLP Analysis

Figure 1 shows ethidium-stained RAPD PCR products amplified from nine unrelated human individuals of diverse ethnicity. A 730-bp product appeared to be polymorphic. To investigate its nature further, this product was isolated and purified from the gel. Labeled 730-bp product was hybridized at high stringency to a Southern blot made from a replicate RAPD agarose gel, showing that this 730-bp fragment is homologous to at least two other RAPD PCR products of various sizes from each of the nine individuals (data not shown).

To assess the extent of its polymorphism, the 730-bp RAPD product was reamplified, labeled, and hybridized to Southern blots containing DNA from four unrelated human individuals digested with nine different restriction enzymes (BIOS RFLP blots). All nine en-

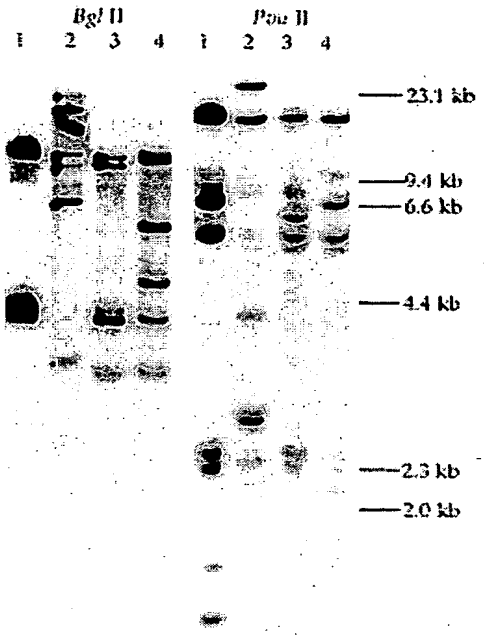

FIG. 2. RFLP patterns of DNAs from four unrelated Caucasian individuals (lanes **1–4**) digested with *Bgl*II (**left**) and *Pvu*II (**right**). The probe was the 730-bp fragment indicated in Fig. 1. Positions of size markers are indicated to the right.

zymes revealed polymorphisms; the *Bgl*II and *Pvu*II results are shown (Fig. 2). Hybridization analysis of DNAs representing a standard CEPH family provided evidence that the alleles detected by the PCR probe are inherited in a Mendelian manner (Fig. 3). The 730-bp probe identified multiple bands from this locus. Haplotype analysis of these allelic bands is beyond the scope of the present work. However, the 4.4-kb fragment is apparently associated with the 2.5-kb band in this family. A short exposure time was required for these blots (less than 3 h).

Sequence of the 730-bp PCR Product

The full sequence of the PCR product is given in Fig. 4A (GenBank Accession No. U22345). The same sequence is organized as six intact and three truncated repeat units arranged in tandem (Fig. 4B). The monomer repeat consensus sequence is 100 bp in length. Sequence similarities between complete monomers are up to 94%. The consensus sequences themselves contain a less conserved direct repeat substructure, and the two halves (1–49 and 50–100) share 78% sequence identity. Thus, there are two levels of repetition within the 730-bp PCR product. BLAST search of the consensus sequences in GenBank and EMBL databases failed to identify any significant matches among published sequences.

Copy Number Estimate

The copy number of the repeat element in the human genome was estimated in a dot-blot assay in which defined amounts of immobilized human DNA were compared with the 730-bp cloned DNA mixed with carrier mouse DNA (Fig. 5). After hybridization with labeled 730-bp PCR product, 5×10^4 copies of human genomic DNA (170 ng) exhibited a hybridization signal equal to that of 10^6 copies of cloned 730-bp PCR product (5.2 pg). This result shows that the haploid human genome contains at least 20 730-bp fragment equivalents of this repeat ($5 \times 10^4 \times 20 = 10^6$). Since a single 730-bp fragment contains six complete and three truncated repeat units (Fig. 4B), our results show that there are a minimum of 140 to 160 copies of the repeat element per haploid human genome (20×7 to $20 \times 8 =$ 140 to 160). This could be an underestimate if there is considerable additional sequence diversity in the repeat family that is not represented in the 730-bp fragment, because all hybridizations were performed at high stringency.

Chromosomal Localization and FISH

The chromosomal localization of the 730-bp PCR product was determined by Southern analysis of a somatic cell hybrid panel (BIOS-MAP SCH/*Hin*dIII panel, BIOS Laboratories). The radiolabeled PCR product hybridized to human-specific bands only in cell lines SM 756 (containing a hamster background plus human chromosomes 6, 7, 13, 14, 19, 20, 21, Y, and partially deleted chromosome 5) and SM 940 (containing hamster background plus human chromosomes 5 and 20). No signals can be detected in other hybrid cell lines containing chromosomes 5, 6, 7, 13, 14, 19, 21, and Y (data not shown). Thus, discordance analysis allowed unequivocal localization of this repetitive sequence to human chromosome 20. This result was also confirmed independently by hybridization to the BIOS monochromosomal panel (data not shown), and long

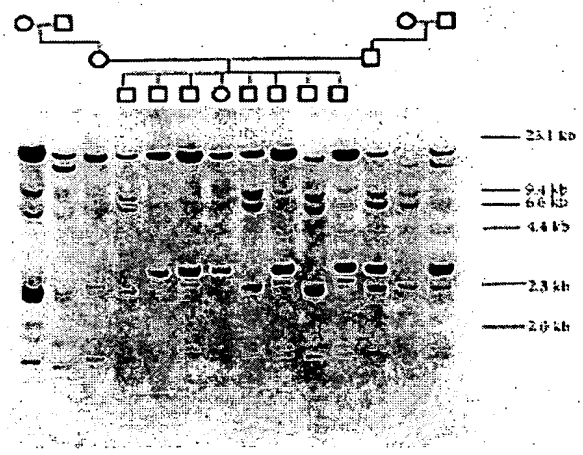

FIG. 3. Pedigree of CEPH Family 1333 aligned with Southern blot showing stable Mendelian inheritance of the novel repetitive element. DNAs (8 μg) were digested with *Pvu*II, electrophoresed in an agarose gel, transferred to a nylon membrane, and hybridized to the 730-bp fragment indicated in Fig. 1. Positions of size markers are given to the right.

```
A
1    ACGACCCACG GAGGAGAGAG GTCTTGTTCT GGTCTGAGTG TCGTGGAGCT GCAGACCCAG GGAGGAGAGG TCTTGTTCTG GTCTGAGGGT CATGGAGCTG
101  CAGACCCAGG GAGGAAGCAA TGTTGTTGTA ATTCGAAGGG TCTGGTAGCT GGCAACTGGG GAGGAGAAGT CTTGTTCGAG TTTGAGGCTG GTGGAGCTGC
201  TGATCAGGGA GGAGCAATTT TGTTGTAGTT CGAGGGTCAG GAAGCTGGTG ACCCAGGGAG GAGAGGTCTT GTTTTAGTTT GAGGGTGGTG GAGCTGCAGA
301  CCCAGGGAGG AGCAATTTTG TTGTAATTCG AGGGTCTGGA AGCTGGAGAC CCAGGGAGGA GAGGTGTTCC GGTTCTAGTT TGAGGGTCGT GGAGCTACAT
401  ACCCAGGGAG GAGCAATCTT GTTGTAATTC GAGGGTGTGG TAGCTGGCAA CCTGGGAGGA GAGGTCTTGT TCTAGTTTGA GGGTCGTGGA GCTGCCGAAC
501  CAGGAGCCAA TTTGGTTGTA ATTGGAGGGT CCCGCAAGCT GGTCGACCCA GGGAGGAGAT GTCTTGTTCT GGTCTGAGGG TCGTGGAGCT ACATACCCAG
601  GGAGGAGCAA TGTTGTTGTA ATTCGAGGGT GTGGTAGCTG GCAACCTGGG GAGGAGAAGT CTTGTTCGAG TTTGAGGCTC GTGGAGCTGC CGAGCAGGGG
701  GGAGCAATTT TGTTGTAGTT CGTGGGTCGT

B          1                                                 50                                                100
CONSENSUS  AGTTTGAGGG TCGTGGAGCT GCAGACCCAG GGAGGAGCAA TTTTGTTGTA ATTCGAGGGT CCTGGAAGCT GGCCGACCCA GGGAGGAGAG GTCTTGTTCT

MODULE#
    1                                                                                      A....... C......... ..........
                                                                                                     /ag
    2      G..C....T. .......... .......... .......AGG .C.....C.G G.CT...... .A....-... .-.A...... .......A-. CAA-------
    3      ---------- ---------- ---------- ---------- .G........ .......... .-...T.... ...A-A-.TG .........A .........G
                                                                  /a
    4      ........C .G........ ..T..-T... .......... .......... G......... .-A....... ..T-...... .......... ........T.
    5      .......... .G........ .......... .......... .......... .......... .-........ ..A-...... .......... ..G.......
                                                                                                              /ccg
    6      .......... .......... A..T...... .......... .C........ .......... G-...T.... ...--.A..T .......... ..........
    7      .......... .......... ..C..A.... .----..... ...G...... ...G...... ..C.C..... ..T....... ........T ..........
                                                 /c
    8      G..C...... .......... A..T...... .......... .G........ .......... G-...T.... ...-A...TG .........A .........G
    9      .........C .......... ..C..G.-.. ..G....... .......... G....T.... .G.
```

FIG. 4. (A) Nucleotide sequence of the 730-bp PCR product. (B) Modular analysis. The PCR product contains six complete and three truncated monomers as represented by each line beneath the consensus sequence. The positions of deletions (−) or insertions (/) are based on best alignment with the consensus sequence. Periods indicate areas of identity to the consensus sequence.

exposure (up to 3 days) revealed no other hybridization targets beside chromosome 20. Fluorescence *in situ* hybridization (FISH) further confirmed this chromosomal localization and established a distinct subchromosomal location of the repeats (Fig. 6). A detectable hybridization signal from the biotin-labeled repetitive element was restricted to the long arm of a group F chromosome (Fig. 6A). In a separate experiment, metaphase chromosomes cohybridized with a mixture of the repetitive element probe, and the biotin-labeled D20Z1 chromosome 20-specific centromere probe (Oncor) (Fig. 6B) demonstrated that the repeats are clustered on the long arm of chromosome 20. Measurements of 10 specifically hybridized chromosomes 20 revealed that the repetitive elements are positioned at 43% of the centromere-to-telomere distance of arm 20q, corresponding to band 20q12 (Fig. 6C).

Analysis of Evolutionary Conservation

Three BIOS "Evo-Blots" (primate, mammalian, and genetic model system) were used to evaluate the evolutionary conservation of this repetitive sequence. Under high stringency, the human repetitive sequence cross-hybridized to sequences from hominoids (chimpanzee, gorilla, orangutan) and Old World monkeys (macaque) only; other primates, such as cotton-top tamarin (*Saguinus oedipus*), failed to yield distinct bands (data not shown). No hybridization signals from other mammals were detected under routine hybridization and washing conditions. The repeat probe revealed polymorphisms among different chimpanzees (Fig. 7).

Genomic Clones

Four positive P1 clones (ICRFP700B0911, ICRFP-700D0363, ICRFP700D1020, and ICRFP700N1820)

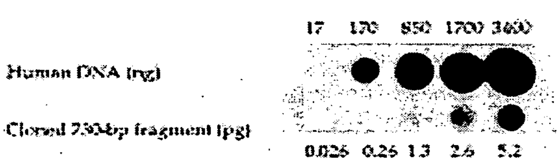

FIG. 5. Dot filter hybridization to estimate copy number of the repetitive sequence. A total of 5×10^4 copies of human genomic DNA (170 ng) exhibited a hybridization signal equal to that of 10^6 copies (5.2 pg) of cloned 730-bp fragment.

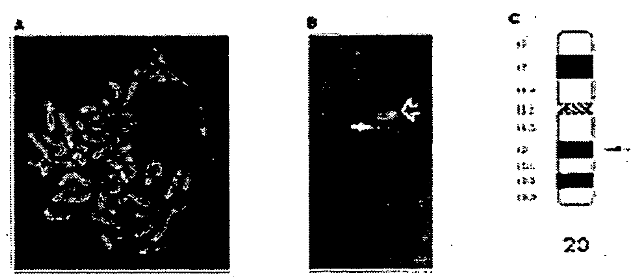

FIG. 6. Chromosomal localization of the repetitive element. (A) Hybridization of cloned, biotin-labeled 730-bp RAPD fragment to human metaphase chromosomes. (B) Same as A except this represents a separate hybridization in which metaphase chromosomes were cohybridized with both the 730-bp fragment (solid arrow) and a biotin-labeled chromosome 20 centromere-specific probe (D20Z1) (open arrow). (C) Idiogram of G-banded chromosome 20, with arrow indicating the chromosomal sublocalization of the repeat cluster.

FIG. 7. Polymorphism in five chimpanzees. DNAs (8 μg each) were digested with *Bgl*II, electrophoresed, transferred to a nylon membrane, hybridized to the ^{32}P-labeled 730-bp fragment, and washed at high stringency.

were identified from the RLDB human P1 library using the 730-bp probe. Southern hybridization patterns of the purified DNA from these P1 clones were compared with total human DNA (data not shown). The hybridization results demonstrated that this relatively small number of clones accounts for most of the repetitive element-positive bands in human genomic DNA. Three of the four P1s are 80% overlapping. Taking into account the fact that average P1 inserts range from 80 to 100 kb, we estimate the domain containing the repetitive sequence to be 200 to 400 kb in size.

DISCUSSION

We have described the structure, sequence, and distribution of a novel human chromosome locus-specific repetitive sequence. RAPD is an efficient method for isolating new repetitive sequences from a genome. The technique was originally developed to facilitate polymorphism studies in species lacking preexisting genetic markers (Williams *et al.*, 1990). Recent studies have shown that RAPD PCR strongly prefers the amplification of repetitive sequences over single-copy sequences in genomic templates (Ayliffe *et al.*, 1994; Tourmente *et al.*, 1994). RAPD requires low-stringency annealing of primers while taking advantage of the greater copy numbers of a repetitive sequence over unique sequences; the repetitive component of genomic DNA is a relatively low complexity template that is preferred during the early stages of RAPD PCR. These sequences exponentially "out-compete" less abundant targets. Amplification of known repeated sequences can be avoided if primers do not anchor within any known repetitive sequences. Thus, RAPD PCR of DNA from somatic cell hybrids and radiation hybrids will be useful in a directed, systematic search for chromosome- and locus-specific repetitive sequences.

The repetitive sequence reported here apparently is restricted to and is highly clustered on human chromosome 20q12. Moyzis *et al.* (1987) reported human chromosome-specific repetitive sequences that are located in heterochromatin at human chromosome positions 9qh and 16qh. However, the repetitive sequences that they characterized are centromere-specific and contain sequences similar to the consensus satellite 2 and 3 sequences. In contrast, the sequence reported here is distinctly noncentromeric yet chromosome-specific. It is located in euchromatin on chromosome 20q12 and bears no homology to α-satellite sequences.

Our FISH result reveals a highly localized cluster of the repeat elements at the cytological level. However, it does not by itself rule out the possibility that additional copies of the repeat element may occupy a more decentralized distribution on chromosome 20; given the small size of the probe, only the aggregate hybridization signal of clustered repeats would be detectable by FISH. To confirm the clustered organization of the repetitive elements, genomic P1 clones were analyzed. Only four positive clones were obtained from a 1.2 × P1 library. Southern hybridization of the purified DNA of representative P1 clones in comparison with total human DNA demonstrated that these clones account for almost all of the repetitive element-positive bands in human genomic DNA. Hence, the repetitive element at 20q12 appears to be highly clustered according to the available evidence.

Our data and those of others (Das *et al.*, 1987; Stallings *et al.*, 1992) suggest that chromosome-specific or locus-specific clusters of repeats may be common. Such repeats have many potential applications. First, as FISH probes, they can provide accurate and rapid identification of individual chromosomes or specific regions of chromosomes because of their genomic reiteration. Since these repeat probes are likely to represent polymorphic sites, they can be used to bridge genetic with cytogenetic maps.

Second, if they are located near tumor-suppressor genes, locus-specific repeat clusters may have important recombinational roles in chromosomal aberrations relevant to cancer. Reported association of the risk of common types of cancers with mutations in HRAS1 minisatellite DNA (Krontiris *et al.*, 1993) suggests that this might be true. Instability of repetitive sequences is well known (Fishel *et al.*, 1994). The repeat element described in this paper maps to a location that is frequently rearranged in acute lymphocytic or nonlymphocytic leukemia, chronic myeloid leukemia, chronic myelomonocytic leukemia, and such myeloproliferative disorders as polycythemia vera, idiopathic myelofibrosis, or refractory anemia (Mitelman, 1991).

A third application concerns forensic identification of individuals. Human-specific polymorphic repetitive elements such as that on 20q12 are urgently needed in forensic science to genotype DNA samples in trace amounts without depending on PCR amplification. The repetitive sequences reported here are inherited in a Mendelian fashion. Polymorphisms were detected by

hybridization with the repeat in Southern analyses of unrelated human individuals of various ethnic groups including Asians, Caucasians, and African-Americans.

ACKNOWLEDGMENTS

We thank Marc Valentine of St. Jude Children's Research Hospital for his excellent technical assistance with FISH, and G. Zehetner, F. Francis, and H. Lehrach of the RLDB for their assistance with the P1 library. This study is partially supported by NIH SBIR Grant R43GM50658-01 to G.R.

REFERENCES

Ayliffe, M. A., Lawrence, G. J., Ellis, J. G., and Pryor, A. J. (1994). Heteroduplex molecules formed between allelic sequences cause nonparental RAPD bands. *Nucleic Acids Res.* **22:** 1632–1636.

Das, H. K., Jackson, C. L., Miller, D. A., Leff, T., and Breslow, J. L. (1987). The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron. *J. Biol. Chem.* **262:** 4787–4793.

Fields, C. A., Grady, D. L., and Moyzis, R. K. (1992). The human THE-LTR(O) and *Mst*II interspersed repeats are subfamilies of a single widely distributed highly variable repeat family. *Genomics* **13:** 431–436.

Fishel, R., Ewel, A., Lee, S., Lescoe, M. K., and Griffith, J. (1994). Binding of mismatched microsatellite DNA sequences by the human MSH_2 protein. *Science* **266:** 1403–1405.

Francis, F., Zehetner, G., Hoglund, M., and Lehrach, H. (1994). Construction and preliminary analysis of the ICRF human P1 library. *Genet. Anal. Tech. Appl.* **11:** 148–157.

Harnden, D. G., Klinger, H. P., Jensen, J. T., and Kaelbling, M. (1985). "An International System for Human Cytogenetic Nomenclature (1985)," Karger, New York.

Hwu, H. R., Roberts, J. W., Davidson, E. H., and Britten, R. J. (1986). Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. *Proc. Natl. Acad. Sci. USA* **83:** 3875–3879.

Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985). Hypervariable "minisatellite" regions in human DNA. *Nature* **314:** 67–73.

Kaplan, D. J., Jurka, J., Solus, J. F., and Duncan, C. H. (1991). Medium reiteration frequency repetitive sequences in the human genome. *Nucleic Acids Res.* **17:** 4731–4738.

Krontiris, T. G., Devlin, B., Karp, D. D., Robert, N. J., and Risch, N. (1993). An association between the risk of cancer and mutation in the HRSA1 minisatellite locus. *N. Engl. J. Med.* **329:** 517–523.

Lane, M. J., Waterbury, P. G., Carroll, W. T., Smardon, A. M., Faldasz, B. D., Peshick, S. M., Mante, S., Huckaby, C. S., Kouri, R. E., Hanlon, D. J., Hahn, P. J., Scalzi, J. M., and Hozier, J. C. (1992). Variation in genomic Alu repeat density as a basis for rapid construction of low-resolution physical mapping of human chromosomes. *Chromosoma* **101:** 349–357.

Lichter, P., Tang, C.-J. C., Call, K., Hermanson, G., Evans, G. A., Housman, D., and Ward, D. C. (1990). High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* **247:** 64–69.

Mermer, B., Colb, M., and Krontiris, T. G. (1987). A family of short, interspersed repeats is associated with tandemly repetitive DNA in the human genome. *Proc. Natl. Acad. Sci. USA* **84:** 3320–3324.

Mitelman, F. (1991). "Catalog of Chromosome Aberrations in Cancer," 4th ed., Wiley-Liss, New York.

Moyzis, R. K., Albright, K. L., Bartholdi, M. F., Cram, L. S., Deaven, L. L., Hildebrand, C. E., Joste, N. E., Longmire, J. L., Meyne, J., and Schwarzacher-Robinson, T. (1987). Human chromosome-specific repetitive DNA sequences: Novel markers for genetic analysis. *Chromosoma* **95:** 375–386.

Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J., Burks, C., Sirotkin, K. M., and Goad, W. B. (1989). The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* **4:** 273–289.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235:** 1616–1622.

Nelson, D. L., Ballabio, A., Victoria, M. F., Pieretti, M., Nies, R. D., Gibbs, R. A., Maley, J. A., Chinault, A. C., Webster, T. D., and Caskey, C. T. (1991). Alu-primed polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: Identification of clones associated with a disease locus. *Proc. Natl. Acad. Sci. USA* **88:** 6157–6161.

Paulson, K. E., Deka, N., Schmid, C. W., Misra, R., Schindler, C. W., Rush, M. G., Kadyk, L., and Leinwand, L. (1985). A transposon-like element in human DNA. *Nature* **316:** 359–361.

Ruano, G., Lewis, M. E., and Kouri, R. E. (1993). Cyclized primer extension: A method for simultaneous DNA radiolabeling and synthesis from templates of unknown sequence. *Anal. Biochem.* **212:** 1–6.

Stallings, R. L., Torney, D. C., Hilderbrand, C. E., Longmire, J. L., Deaven, L. L., Jett, J. H., Doggett, N. A., and Moyzis, R. K. (1990). Physical mapping of human chromosomes by repetitive sequence fingerprinting. *Proc. Natl. Acad. Sci. USA* **87:** 6218–6222.

Stallings, R. L., Doggett, N. A., Okumura, K., and Ward, D. C. (1992). Chromosome 16-specific repetitive DNA sequences that map to chromosomal regions known to undergo breakage/rearrangement in leukemia cells. *Genomics* **13:** 332–338.

Tourmente, S., Deragon, J. M., Lafleuriel, J., Tutois, S., Pelissier, T., Cuvillier, C., Espagnol, M. C., and Picard, G. (1994). Characterization of minisatellites in *Arabidopsis thaliana* with sequence similarity to the human minisatellite core sequence. *Nucleic Acids Res.* **22:** 3317–3321.

Waye, J. S., and Willard, H. F. (1985). Chromosome-specific alpha satellite DNA: Nucleotide sequence analysis of the 2.0 kilobasepair repeat from the human X chromosome. *Nucleic Acids Res.* **13:** 2731–2743.

Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44:** 388–396.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafiski, J. A., and Tingey, S. V. (1990). DNA polymorphism amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* **18:** 6531–6535.

Zehetner, G., and Lehrach, H. (1994). The reference library system—Sharing biological material and experimental data. *Nature* **367:** 489–491.

Zucchi, I., and Schlessinger, D. (1992). Distribution of moderately repetitive sequences pTR5 and LF1 in Xq24–q28 human DNA and their use in assembling YAC contigs. *Genomics* **12:** 264–275.